DOCUMENT RESUME

ED 414 312 TM 027 788

AUTHOR Fenton, Ray; Straugh, Tom; Stofflet, Fred
TITLE Alaska Writing Assessment - 1997: Preliminary Technical Report.
PUB DATE 1997-09-26
NOTE 35p.; Paper presented at the "Standards in Action Forum" of the Alaska State Department of Education (Anchorage, AK, September 26, 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Secondary Education; Evaluation Methods; *Program Implementation; Resource Allocation; *Scoring; *State Programs; *Test Construction; Test Items; Test Results; *Testing Programs; Writing Achievement; *Writing Tests
IDENTIFIERS *Alaska Writing Assessment

ABSTRACT

Writing assessment began in Alaska in the 1970s, and the Alaska Writing Assessment (AWA) that was piloted in 1997 built on previous efforts. The 1997 AWA involved more than 20,000 students in grades 5, 7, and 10 from 43 school districts, and the mandatory assessment planned for 1998 will include approximately 28,000 students. This review of the 1997 AWA will help in planning for the 1998 AWA. The 1997 AWA computer file kept records on more than 16,000 students in the 3 grades (grades 5, 7, and 10). Students were scored using a six-trait analytical writing assessment scoring guide developed by the Northwest Regional Educational Laboratory that provides generic descriptions of characteristics of performances. Overall averages ranged from a low of 3.08 in "Organization" to a high of 3.28 in "Ideas and Content," and the overall average score across all areas was 3.22 with a standard deviation of 0.71. A review of the AWA for 1997 shows that the promise of a true statewide writing assessment was realized. The 1997 AWA extended a small-scale voluntary program to a program that demonstrated the capacity to conduct statewide writing assessment. The problems discovered were substantial, but can be solved. Additional human and financial support must be provided to organize and conduct an assessment that will collect papers from so many students and return valid scores. The quality of scoring must be improved to assure that scores are valid for the intended uses of rating the performance of individual students, providing feedback to teachers and schools, and allowing comparison of groups for local, state, and federal program evaluation. In addition, the process of developing prompts must be regularized so that prompts can be tested with Alaska students. The six-trait scoring guide is attached. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 414 312

Alaska Writing Assessment - 1997 Preliminary Technical Report

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Ray Fenton
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Ray Fenton
Tom Straugh
Fred Stofflet

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Assessment and Evaluation Department[0]
Anchorage School District

TM027785

A paper prepared for presentation to the "Standards in Action Forum"
Alaska State Department of Education
Anchorage, AK
September 26, 1997

---

[0] **Assessment and Evaluation Department, Anchorage School District, P.O. Box 196614, Anchorage, AK 99519. Phone: (907) 269-2211 Fax: (907) 269-2222 e-mail: fenton_ray@msmail.asd.k12.ak.us**

BEST COPY AVAILABLE

Alaska State Writing Assessment 1997[1]

Promise and Problems

Writing assessment started in Alaska in the 1970s following a model advocated by the Bay Area Writing Project. The State of Alaska has supported an Alaska Writing Project and the assessment of writing as part of instruction since the early 1980s. The original Alaska writing assessment developed as a voluntary effort through the Alaska Writing Project and came to be known as the Alaska Voluntary Direct Writing Assessment. It was organized and encouraged as part of the Alaska Writing Project and school districts were able to submit writing if they agreed to provide a teacher to score papers. A statewide prompt committee would select prompts for two grades and participation varied from year to year depending on funding and the emphasis on writing at individual school districts.

The Alaska Commissioner of Education, Shirley Holloway, mandated that the 1997 writing assessment be extended to all students in grades 5, 7, and 10 as part of the Alaska Statewide Student Assessment. Under the state regulation, every student except those students who were not fluent in English and had not been in an English speaking school for three years including the year of the assessment or students in special education who were specifically excluded from testing is to be assessed.

After protests from a few school districts which have developed their own effective writing programs include active writing assessment programs, the Commissioner allowed the 1997 year to serve as a pilot year for Alaska Writing Assessment (AWA). All Districts were encouraged to participate, but participation was not required. All districts will be required to participate in 1998.

The State of Alaska Administrator for Standards and Assessment, Dennis McCrea, was charged with the implementation of the Alaska Writing Assessment. He directed pilot project in 1997 and provided an overview of the plans for the 1998 writing Assessment to District Superintendents and District Assessment Administrators in August 1997. Highlights of that memo include:

- More than 20,000 students in grades 5, 7, and 10 from 43 school districts participated in the AWA.
- The 1998 assessment is mandatory and will include approximately 28,000 students.
- Writing will be evaluated for ideas, organization, voice, word choice, sentence fluency and conventions.
- Classroom teachers will administer the 1998 writing assessment during three class periods (150 minutes) during the period of January 12-30.
- Districts will be expected to provide teachers to score papers at two or perhaps three locations including Anchorage March 3-6 and Fairbanks March 10-13. Participants may opt for participation in an additional class on Teaching Process Writing in the Classroom on the day following the paper scoring.
- Student papers and scores are to be returned to students and teachers prior to the end of the school year.

---

[1] Analysis of Alaska Writing Assessment data was done with the permission of Dr. Dennis McCrea, Administrator for Standards and Assessment of the Alaska Department of Education. A formal report on the 1997 AWA performance is available from the Alaska State Department of Education, 801 West 10th Street, Suite 200, Juneau, AK 99801. Phone (907)465-2830.

BEST COPY AVAILABLE

Because the intention for 1998 is to follow the procedures of the 1997 large scale pilot, it is important that we consider what happened in 1997, good and bad; and consider what we might learn from the experience.

The 1997 Assessment

The 1997 Assessment was planned late in 1996 and early in 1997. Because of changes in State Department of Education personnel, real planning did not start to take place until late in 1996. There were a number of major decisions made early in the process about the writing which would be expected of students and the procedures for assessing writing.

- There would two types of writing prompts - Informative and Persuasive. Students would be randomly asked to provide samples of one or the other type of writing.
- There would be two prompts for each type of writing. Students would be free to choose one or the other of two prompts provided on their blue book.
- Prompts would be selected from prompts which had been used in prior years in the State of Oregon.
- Prompts would not be pre-tested with Alaska students due to the limited time available.
- Judy Arter and Vicki Spandel of Northwest Regional Educational Laboratory would train the scorers and manage the process of scoring papers.
- Scoring would be based on rubrics developed for past writing assessments and sample papers provided by NWREL rather than materials generated specifically for the 1997 assessment.
- Papers which were not scored in Alaska would be scored at NWREL in Portland.
- Northwest Educational Laboratory and the Anchorage School District would work to scan student surveys, score reports, and generate student reports.
- The Alaska State Department of Education would manage the distribution of materials, collection of materials, preparation of materials for scoring, the return of papers and results to school districts, and the development of summary reports on the success of students in meeting.

The plan for the implementation and operation of the 1998 Alaska Writing Assessment is to be much the same as the 1997 pilot. Because of this, it is important that we take the time to examine the successes and failures of the pilot to focus and improve the 1998 AWA.

Ray Fenton and Tom Straugh of the Assessment and Evaluation Department of the Anchorage School District played an active role in the planning and implementation of the Alaska Writing Assessment. Dr. Fenton served on the Alaska State Department of Education Committee which recommended that writing assessment be the first extension of statewide testing beyond norm referenced testing. Dr. Straugh participated in the planning discussions and was on the prompt selection committee. Drs. Fenton and Straugh worked with Northwest Laboratory staff in the collection and scanning of data and with the State Department of Education in preparing reports for schools, teachers, and students. Statistical analyses presented in this report were conducted by Dr. Fenton using SPSS-X and a data base developed from papers scored in Anchorage and Portland, OR.

A series of research questions are examined relative to the 1997 Alaska Writing Assessment were developed through conversations among Dr. Fenton, Dr. Straugh, Dr. Stofflet, and Dr. McCrea. The questions are answered based on the experience of the authors with the 1997 assessment and analysis of the data base. A series of conclusions are reached and suggestions made for improvement of the 1998 AWA based on the analysis and the experience of the authors with the 1997 Alaska Writing Assessment.

Research Questions

What were the scores Alaska students received on writing?

To what extent did those who graded papers agree with each other?

Did the type of writing selected affect the scores?

Did the prompt selected affect the scores?

What can examination of the scores tell us about the efficacy of the six trait model.

What did survey items tell us about writing instruction in Alaska?

What did we learn about the procedures used for scoring?

What did we learn about the procedures used for distribution, collection, and reporting?

What were the scores Alaska students received on writing?

The computer file contained complete records for over 16,000 students including 6,215 students in grade 5, another 5,434 students in grade 7, and another 4,885 students in grade 10. Not all of the files were complete as some records were missing information on grade, prompt selected, and survey items.

Students were scored using a six-trait analytical writing assessment scoring guide developed by Northwest Regional Educational Laboratory (Appendix A). The guide provides generic descriptions of the characteristics of performances which might be rated 1 (Beginning), 2 (Emerging), 3 (Developing), 4 (Closing In), or 5 (Writing With Purpose & Confidence). The general characteristics of papers which have a score of 1, 3, and 5 are described for each trait: Organization, Voice, Word Choice, Sentence Fluency, Conventions. The approach used by NWREL has teachers score a generic set of papers which are not on the topics being scored and "certifying" a reader as a scorer if they are able to generally reach agreement with the score on the sample set of papers.

Papers were distributed and scored without regard to grade level or prompt. Each grade level group of papers contained writing from one grade and either informative or persuasive writing. Individual students choose the prompt so every Rater was reading papers on 12 different prompts drawn from 3 grades.

If a score of 1 or 2 is considered as "Below Expectation" and a score of 3, 4 or 5 is considered to be at or above expectation, most Alaska students are at or above expectation in all of the traits of writing. The scores for students who had scores in all areas are presented below.

Alaska Writing Assessment 1997
Overall Scores

Trait	*Below Expectation*	*Above Expectation*	*Total*
Ideas	2,816	8,639	11,483
	24.5%	75.2%	100%
Organization	3,797	7,658	11,483
	33.1%	66.8%	100%
Voice	1,846	9,612	11,483
	16.1%	83.7%	100%
Word Choice	2,447	9,011	11,483
	21.3%	78.5%	100%
Sentence Fluency	2,873	8,585	11,483
	25%	74.8%	100%
Conventions	2,971	8,486	11,483
	25.9%	73.9%	100%

When differences are considered by grade, the performance is generally similar with about three out of four students at or above a scores of three. There is a progression of scores with more students found at or above expectation at grade 10 than 7 and more at grade 7 than at grade 5 suggesting that more students come up to the expectations of teachers at the years progress. The differences are statistically significant and support the idea that there was an ability to discriminate

performance at the different levels though there is no metric which allows a comparison of the performance at the various levels to get a sense of how much better average performance might have been at grade 7 or grade 10.

The overall averages ranged from a low of 3.08 in Organization to a high of 3.28 in Ideas and Content. The standard deviations associated with average scores were fairly large ranging from .90 for Ideas and Content to .71 for word choice. The overall average score across all of the trait areas was 3.22 with a standard deviation of .71 suggesting that a difference of half a score point would not be unusual.

Alaska Writing Assessment 1997
Proficiency by Grade Level

Trait	*At or Above Expectation Grade 5*	*At or Above Expectation Grade 7*	*At or Above Expectation Grade 10*
Ideas	74.3%	72.0%	84.8%
Organization	59.1%	64.7%	78.1%
Voice	80.1%	81.8%	90.2%
Word Choice	75.7%	72.9%	87.9%
Sentence Fluency	69.0%	71.5%	85.3%
Conventions	68.7%	71.4%	83.0%

To what extent did those who graded papers agree with each other?

Validity it testing is a simple concept but difficult to prove in a given case. The best we can do is make arguments for the validity of scores. To be valid, a score must be useful for its intended purpose. In the case of the Alaska State Writing Assessment, there are multiple purposes and each needs to be considered.

- Is a score a valid indicator of the desired quality of writing? Is the criteria correct?
- Is a score valid as an indicator of how an individual or a group of students are doing as writers? Is the score a reliable indicator?
- Is the score valid for making comparisons of individuals, groups, programs, schools?
- Is the score valid for making comparisons from time to time to assess progress?

Reliability of scores is the first prerequisite of validity. If a performance can not be consistently judged and consistently assigned the same score, the score can not said to represent the performance and there can be no basis for comparison of scores.

Writing assessment assures reliability of scoring by having more than one individual score a paper or set of papers and then examining the level of agreement among scorers and between scorers and a fixed standard. The initial comparison is generally between a Rater and a set of papers that have been selected to represent the various score levels on the traits being examined. In the case of the Alaska Writing Assessment, all readers were "certified" based on a demonstration that they were able to give scores to an initial set of papers similar to the expected scores upon conclusion of the training conducted by NWREL. No record was kept of the agreement of raters on the sample papers.

From time to time, teachers were asked to score sample papers to "recalibrate" during the AWA reading. Individual sample papers were presented and teachers were asked to report on an "honor system" if their scores disagreed with the scores on the sample papers. Sample papers were not drawn from the topics being scored and were not presented on each of the prompts at each of the grade levels. No record was collected as to the extent of continued agreement between readers and a sample set of papers.

Individual raters were assigned code numbers and asked to report their numbers on each of the sheets that were scored. Under usual circumstances this would have allowed an examination of each Rater pair to see if there were consistent agreements or disagreements among raters and to determine if some individuals were consistently high or consistently low in the scores that they assigned. Unfortunately, AWA scoring took place at two locations in Anchorage and raters at the two locations were given the same numbers. There were also some raters who were assigned more than one number. This confusion makes a pair wise examination of Rater agreement impossible. Anecdotal reports from raters to the authors suggested that there were differences with some "hard" and some "easy" graders.

Each paper was scored by two individuals and it was possible to examine the relation of scores without regard to individual raters and Rater pairs.

Alaska Writing Assessment 1997
Rater Agreement by Trait

Trait	Agreement	Off 1	Off 2	Off 3	Off 4
Ideas	48%	45%	6%	1%	-
Organization	46%	47%	6%	1%	-
Voice	48%	46%	5%	-	-
Word Choice	54%	42%	4%	-	-
Sentence Fluency	49%	47%	5%	-	-
Conventions	48%	47%	5%	-	-

The pattern of Rater agreement by grade is similar without significant differences in levels of agreement among the grades. The patterns are also similar by mode of writing and by prompt suggesting that the two raters gave the same grade about 50% of the time and were within one point about 90% of the time.

The highest levels of Rater agreement are with the lowest and highest scores. The lowest level of agreement is with the score of three where the second reader who disagrees is slightly more likely to give a score of 4 than a score of 2. There is a slight bias toward higher scores as is suggested by the overall average scores of slightly above 3 in all areas.

The low rater agreement may be due in part to the training method used. Examination of the method used by NWREL for training teachers indicated that after a review to the traits and typical papers individuals were asked to score a set of 18 generic papers to qualify as readers rather than a set of papers specific to an individual type of writing, prompt, and grade level. There was not a cut score set to disqualify individuals and the readers were asked to grade their own work. Eighteen papers were read and individuals scored an agreement with the score on the criteria paper a "2" and a score within one point a "1" for each of the six traits. A total agreement score would be 216. A score 75% was considered good and a score of 65% was considered acceptable (Attachment 2).

This level of agreement is about equal to what Ellis Page reported for Read American when untrained readers were asked to grade papers using standard rubrics with samples of typical student work. Writing assessments in other states and those conducted by Educational Testing Service have been able to reach 90% agreement with a careful training program, elimination of readers who are unable to consistently grade papers relative to the standard, and regular review of reader performance during the scoring sessions.

Did the type of writing selected affect the scores?

Yes, there were significant differences between prompts related to informing or describing and persuading. The students who were attempting persuasive writing did not score as well as their grade level peers who were writing to inform. The pattern of higher performance in each area at each grade level generally holds up for each of the types of writing.

Alaska Writing Assessment 1997
Proficiency by Type of Writing
by Grade Level

Trait	*Grade 5 Inform*	*Grade 5 Persuade*	*Grade 7 Inform*	*Grade 7 Persuade*	*Grade 10 Inform*	*Grade 10 Persuade*
Ideas	74.3%	66.0%	74.5%	69.5%	87.5%	82.3%
Organization	60.9%	57.3%	65.8%	63.6%	81.6%	74.8%
Voice	81.0%	79.1%	82.7%	80.9%	90.6%	89.9%
Word Choice	79.5%	71.5%	75.7%	70.2%	89.2%	86.6%
Sentence Fluency	71.9%	65.9%	73.6%	69.5%	88.0%	82.6%
Conventions	69.0%	68.3%	72.4%	70.3%	84.4%	81.8%

Did the prompt selected affect the scores?

Yes, there are a number of cases where one prompt has significantly higher scores when the two available prompts are examined. The differences in scores appear to be generally consistent by trait area with the higher or lower scoring prompt being consistently higher or lower. There is no way to tell from the data if one prompt was "harder" than the other resulting in differences in content and length of the written product or if one prompt was more "attractive" to the scorer than the other with the student getting a more positive or negative rating because of the choice of subject.

Systematic differences in prompt selection in different classes or districts could result in substantial differences in class or district scores.

Alaska Writing Assessment 1997
Proficiency Informative Writing by Prompt Selected
by Grade Level

Trait	*Grade 5 Prompt 1*	*Grade 5 Prompt 2*	*Grade 7 Prompt 1*	*Grade 7 Prompt 2*	*Grade 10 Prompt 1*	*Grade 10 Prompt 2*
Ideas	77.3%	70.7%	74.7%	74.4%	80.9%	97.7%
Organization	64.8%	56.2%	68.5%	64.1%	73.9%	86.4%
Voice	80.7%	81.3%	83.9%	81.9%	81.8%	96.2%
Word Choice	82.7%	75.8%	74.5%	76.4%	87.1%	90.6%
Sentence Fluency	74.0%	69.5%	74.0%	73.3%	86.2%	89.2%
Conventions	70.3%	67.6%	69.7%	74.1%	83.0%	85.3%

Alaska Writing Assessment 1997
Proficiency Persuasive Writing by Prompt Selected
by Grade Level

Trait	*Grade 5 Prompt 1*	*Grade 5 Prompt 2*	*Grade 7 Prompt 1*	*Grade 7 Prompt 2*	*Grade 10 Prompt 1*	*Grade 10 Prompt 2*
Ideas	68.3%	63.7%	68.9%	69.9%	81.5%	82.8%
Organization	57.8%	56.7%	65.7%	62.0%	73.3%	75.8%
Voice	79.6%	78.7%	82.1%	79.9%	89.2%	90.4%
Word Choice	74.4%	68.5%	70.9%	69.6%	85.6%	87.3%
Sentence Fluency	67.2%	64.6%	70.4%	68.9%	80.9%	83.3%
Conventions	68.1%	68.5%	72.4%	68.8%	80.6%	82.6%

The differences in prompts are more marked in informative writing than in persuasive writing.

What did survey items tell us about writing instruction in Alaska?

A substantial number of classroom groups did not have completed surveys. It appears that some students were instructed not to complete the informational survey which includes items on the amount and type of writing instruction provided. Of the approximately 16,000 writing samples completed, survey information was not provided for over 2,000 student and as many as 3,500 students did not respond to some individual items related to grades. The consistent omission of survey information from some schools and districts raises a question about the extent to which survey information may be generalized as representative of writing instruction in Alaska.

There are significant relationships between the instruction reported and student performance. In general, the more writing the higher the average score on all of the traits and the more writing activities present the higher the average score. The percent of positive responses is included for each survey item. There are notable and significant differences by grade with older students writing more and not as likely to consider themselves a good writer.

Secondary students are less likely to report that they have formal instruction in spelling and the mechanics of writing on a regular basis though they spend more time writing.

Because 1997 is to be the first year of a true statewide writing assessment, the survey data from 1997 will serve as a baseline for a description and comparison of programs. The data from the 1996 assessment may serve as a basis for comparison, but it can not serve as a baseline due to the selective participation of schools and classrooms in providing information. When there is more complete data, it would be worthwhile to build profiles of more successful writing programs to serve as models of effective instruction in writing.

BEST COPY AVAILABLE

Alaska Writing Assessment
Student Survey Items

Item	Percent "YES"
Learning to write is important.	75%
I am a good writer.	48%
In class writing time of an hour or less a week.	43%
Home writing time of an hour or less a week.	49%
Easy to find a computer to use for writing at school.	65%
Have a computer at home for writing.	68%
Writing is graded based on multiple choice tests.	21%
Writing is graded based on short answer tests.	28%
Writing is graded based on long essays.	31%
Writing is graded based on individual projects, journals, presentations.	54%
Writing is graded based on group products, presentations, portfolios.	36%
Don't know the basis for my writing grades.	24%
We often make outlines.	42%
We often define the purpose of our writing.	32%
We often define the audience.	19%
We often go to sources other than the textbook.	46%
We often write more than one draft of a paper.	54%
We often choose the topic on which we write.	48%
We often edit our own writing for content and style.	43%
We often exchange papers for editing.	47%
We often discuss writing with our teacher.	44%
We spend time on mechanics every day or almost every day.	58%
We work on writing in pairs or in groups at least 1 or 2 times a week.	49%
When we are graded, ideas and content are important.	87%
When we are graded, organization is important.	88%
When we are graded, voice is important.	82%
When we are graded, word choice is important.	83%
When we are graded, fluency is important.	86%
When we are graded, conventions is important.	90%
We do portfolios.	53%
We keep journals.	54%
We publish our best work.	49%
We use rubrics to asses our own work.	44%
We do speaking and listening activities.	77%
We give oral presentations.	68%
Class size less than 25.	44%

What does statistical analysis of the data tell us relative to the six trait model?

Ratings of individual traits are highly correlated. That is, if an individual is likely to be rated high or low on one trait is also likely that the individual will be rated high or low on other traits. For example, the overall correlation of ratings with Ideas and Content are: Organization, r = .75; Voice, r = .62; Word Choice, r = .64; Sentence Fluency, r = .59; and Conventions, r = .48.

The pattern of correlation is fairly consistent across grades, types of writing, and prompts. Either Ideas and Content or Organization account for most of the variation in scores and the correlations are always statistically significant. This lack of independence of traits is sometimes called a "Halo Effect" where the overall quality of writing affects all of the scores. The score which is least affected by the halo effect is Conventions.

Regression analysis suggests that the strongest predictors of the overall score derived from simple addition of the six traits are Ideas and Content, Fluency, and Conventions. More than 80% of the overall variation can be explained with these three scores.

Because each paper was independently scored by two raters, it was possible to do two independent factor analytic studies to see if there was an underlying pattern within the six traits. Each of the independent factor analytic studies came to the same conclusion. The bulk of the variation in scores could be accounted for in a single factor (69% of the variance in study 1, 72% of the variance in study 2) which is made up of two traits: Ideas and Content, and Organization. A second factor adds an explanation for an additional 10% of the variation with another two traits: Sentence Fluency and Conventions. A third factor explains another seven percent or so of the variation: Voice and Word Choice.

If a decision were to be made on scoring writing based on the statistical features of the six trait model writing assessment, it would be to reduce the factors scored to two or three. The two trait model would suggest a holistic scoring for content, organization, and fluency with a separate scoring for spelling and writing conventions. The three trait model would suggest one scoring for ideas, content, and organization; one scoring for sentence fluency and conventions, and a third scoring for voice and word choice.

If a revised model were used to classify students as below basic, basic, or advanced as writers using the overall variation in the six trait scoring; students would be properly classified more than 90% of the time using the two or three trait model.

Of course, the argument in favor of the six trait model has been centered on instruction. The belief is that teachers are able to organize instruction around the six traits and use of the six traits in the assessment of writing encourages students to improve in each of the areas.

It would take exploration of actual practice in writing instruction in Alaska to ascertain if instruction and grading of writing is actually build around the six traits of the writing instruction. The Alaska Statewide Assessment based on the use of standardized test scores contracted for an independent assessment of the efficacy and utility of the standardized test scores to assess and improve instructional programs. There has been no similar effort undertaken with Writing Assessment or the more recent additions to the mandated Alaska Statewide Student Assessment Program: NAEP, reading assessments, and the proposed new national assessment.

BEST COPY AVAILABLE

What did we learn about the procedures used for scoring?

More training and more careful monitoring of readers is necessary. A reader agreement rate of 50% is too low to be acceptable even if the ratings are within one point 90% of the time. There are a number of things which might be done to improve the precision of the scoring. Some of the things which should be considered include:

Training and Monitoring Scoring to Assure Validity

- Using writing which is derived from the actual prompts under conditions similar to those of the actual assessment to establish benchmark papers to be used in training and in the verification that those who are scoring actually use the same standards for scoring each paper.
- Set an absolute criteria for the acceptance of readers and reject individuals as scorers who can not score using the standard.
- Conduct a running examination of Rater agreement while scoring is taking place to identify individuals who are too "hard" or "easy" in marking papers.

Organization of Scoring to Assure Fairness and Equity

- Score only one prompt at one grade level at a time.
- Randomize papers so that readers will not identify individual schools and districts as they score.
- Randomize papers so rural and urban papers are mixed.
- Randomize papers across scoring sites.

Increased Participation in Scoring By Teachers

- Teachers should score papers at or near the grade level they teach.
- Teachers should be drawn from the districts which provide papers in the proportion that papers are provided.
- The ratio of teachers should be based on a reasonable expectation for the number of papers to be read in a day.

What did we learn about the procedures used for distribution, collection, and reporting?

- District superintendents and assessment administrators need advance warning to schedule writing activities and assure that all understand the instructions and need for standardization of the process.
- Training should be provided to district assessment administrators to assure that training of teachers in assessment, application of rules related to bilingual and special education students, and all coding of teacher, school, grade, and student information is accurate and complete prior to submission of papers meets the requirements set by the state.
- Materials need to be distributed directly to schools or to districts in enough time to assure the distribution of materials to rural sites, checking of materials, and time to order additional materials as needed.
- Staff should be available to respond to district, teacher, and parent questions about the writing assessment, procedures, decisions about inclusion and exclusion of students, and requests for additional materials.
- Labels and shipping labels should be available for the return of materials to the state. A common carrier such as Federal Express or Airborne Express should be used to assure that individual shipments of student test papers can be traced and lost packages located.
- Staff should be available when papers are returned to check each school, class, and student paper to assure that coding is complete. When it is not complete, the school should be contacted to gather the correct information.
- Groups of papers should be prepared in advance for scoring sessions. Papers should be randomized across sites and mixed so that individual readers will not recognize a group of papers as "urban" or "rural" and shift the scoring standard.
- Analysis of ratings should take place as scoring is taking place to identify raters who are consistently "too hard" or "too easy" to identify groups of papers for rescoring and raters who need retraining.
- Scanning and data collection should take place in one location. This will reduce the chance of handling problems and increase the likelihood that a school, teacher, and students will have papers returned prior to the end of the school year.
- Summary reports of performance relative to the expected performance should be distributed to school districts in time for use with the State School Report Card.
- Data should be provided to school districts and Title I, Migrant Education, Bilingual Education, coordinators in a format which will allow the tracking of participating students and inclusion of information in State and Federal required reports of program success.

BEST COPY AVAILABLE

Conclusions and Recommendations

The promise of a true statewide writing assessment was realized.

The 1997 Alaska Writing Assessment was a substantial undertaking which extended a small scale voluntary program with occasional participation by some classrooms in some school districts to a true statewide pilot program which demonstrated the capacity to conduct statewide writing assessment.

The papers from approximately 20,000 students spread across 47 school districts were collected and scored.

Reports were generated and returned with scored papers to students and teachers throughout the state.

The problems discovered were substantial but are solvable.

Additional human and financial support must be provided to organize and conduct an assessment which will collect papers from 30,000 students and return valid scores to parents, teachers, and schools.

Quality of scoring must be improved to assure that the scores are valid for the intended uses of rating the performance of individual students, providing feedback to teachers and schools on the success of educational programs in meeting state goals, and allowing comparison of groups for local, state, and federal program evaluation.

- Training and monitoring of those who score must be improved.
- Teachers should score the papers from the grade level with which they are familiar and not be asked to score multiple prompts, different types of writing, and multiple grade levels as the same time.
- The process for handling and coding papers must be improved.
- The process for analysis of results and preparation of results must be improved.

The process of developing prompts, testing prompts, and developing materials for training in scoring must be regularized and done so as to allow trials of prompts with Alaska students prior to writing.

The system for the analysis and reporting of results should matched to a similar effort to train teachers in effective methods of writing instruction and demonstrate that good instruction improves performance.

There should be an independent evaluation of the accuracy and efficacy of the State Writing Assessment with a focus on the utility of the State Writing Assessment as a tool to improve writing instruction.

Six-Trait Analytical Writing Assessment Scoring Guide

- Ideas and Content (Development)
- Organization
- Voice
- Word Choice
- Sentence Fluency
- Conventions

Northwest Regional Educational Laboratory
101 SW Main Street, Portland, Oregon 97204

Excerpted from Creating Writers, 1997. Vicki Spandel and Richard J. Stiggins. New York: Addison-Wesley/Longman Publishers, 1-800-552-2259. The authors gratefully acknowledge the contributions of the Analytical Writing Assessment Committee, a group of 17 teachers from Beaverton, Oregon, who developed in 1984 the original scoring guide upon which this one is based; and in addition, the invaluable insights and reflections offered by countless teachers of writing at all grade levels across the country who have taken the six-trait model into their classrooms and turned it into a powerful revision tool.

BEST COPY AVAILABLE

What the Scores Mean

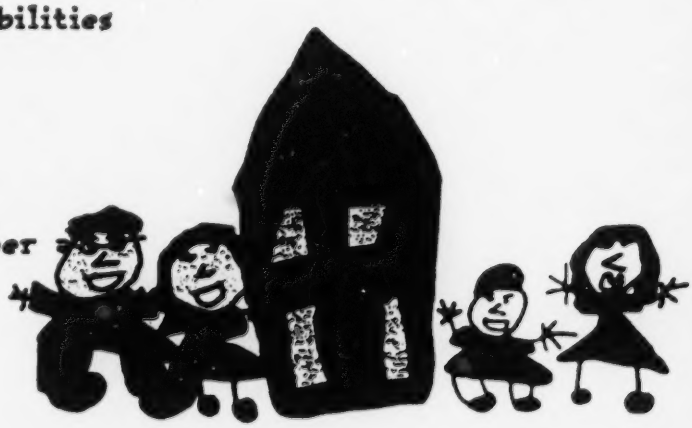
Level 5
Writing With Purpose & Confidence
Making it individual
Fine tuning
Creating writing that speaks to an audience

Level 4
Closing In
Revising with purpose
Adding detail
Feeling it come together
Expanding the vision

Level 3
Developing
Taking control
Acting on the possibilities
Getting a solid foothold
Knowing where it's all headed

Level 2
Emerging
Making discoveries
Creating some possibilities

Level 1
Beginning
Searching, exploring
Getting something on paper



BEST COPY AVAILABLE

IDEA DEVELOPMENT

Ideas are the heart of the message, the main thesis, impression, or story line of the piece, together with the documented support, elaboration, anecdotes, images, or carefully selected details that build understanding or hold a reader's attention.

5 *The paper is clear, focused, purposeful, and enhanced by significant detail that captures a reader's interest.*

- The paper creates a vivid impression, makes a clear point, or tells a whole story, without ever bogging the reader down in trivia.
- Thoughts are clearly expressed and directly relevant to a key issue, theme, or story line.
- The writer selectively and purposefully uses knowledge, experience, examples and/or anecdotes to make the topic both understandable and interesting.
- Quality details consistently inform, surprise, or delight the reader—or just expand his or her thinking.

3 *The writer has made a solid beginning in defining a key issue, making a point, creating an impression, or sketching out a story line. More focus and detail will breathe life into this writing.*

- It is easy to see where the writer is headed, even if some telling details are needed to complete the picture.
- The reader can grasp the big picture but yearns for more specific elaboration.
- General observations still outweigh specifics.
- There may be too much information; it would help if the writer would be more selective.
- As a whole, the piece hangs together and makes a clear general statement or tells a recountable story.

1 *The writing is sketchy or loosely focused. The reader must make inferences in order to grasp the point or piece together the story. The writing reflects more than one of these problems—*

- The writer still needs to clarify the topic.
- The reader often feels information is limited, unclear, or simply a loose collection of facts or details that, as yet, do not add up to a coherent whole.
- It is hard to identify the main theme or story line.
- Everything seems as important as everything else.

BEST COPY AVAILABLE

ORGANIZATION

Organization is the internal structure of the piece. It is both skeleton and glue. Strong organization begins with a purposeful, engaging lead and wraps up with a thought-provoking close. In between, the writer takes care to link each detail or new development to a larger picture, building to a turning point or key revelation and always including strong transitions that form a kind of safety net for the reader, who never feels lost.

5 *The order, presentation, or internal structure of the piece is compelling and moves the reader purposefully through the text.*

- The organization serves to showcase or enhance the central theme or story line.
- Details seem to fit right where they are placed, though the order is often enlivened by a surprise or two.
- An inviting lead draws the reader in; a satisfying conclusion ties up loose ends and leaves the reader with something to think about.
- Pacing feels natural and effective; the writer knows just when to linger over details and when to get moving.
- Organization flows so smoothly the reader does not need to think about it.
- The entire piece seems to have a strong sense of direction and balance. Main ideas or high points stand out clearly.

3 *The organizational structure guides the reader through the text without undue confusion.*

- Sequencing seems reasonably appropriate, given the main theme or story line.
- Placement of details seems workable though not always deft.
- Predictable moments or developments outweigh surprises or discoveries.
- The introduction and conclusion are recognizable and functional.
- Transitions are usually present but sometimes reinforce obvious connections.
- Structure is sometimes so dominant it is hard for the reader to focus on the ideas or voice.
- The piece has a developing sense of balance; the writer is zeroing in on what is most important but does not yet build to that point with a strong sense of momentum.

1 *Ideas, details, or events seem loosely strung together. The reader struggles to discover a clear direction or purpose. The writing reflects more than one of these problems—*

- There is as yet no identifiable structure to move the reader from point to point.
- No real lead sets up what follows.
- No real conclusion wraps things up.
- Missing or unclear transitions force the reader to make giant leaps.
- Sequencing feels more random than purposeful, often leaving the reader with a disquieting sense of being adrift.
- The writing does not build to a high point or turning point.

BEST COPY AVAILABLE

VOICE

Voice is the presence of the writer on the page. When the writer's passion for the topic and concern for the audience are strong, the text virtually dances with life and energy, and the reader feels a strong connection to both writing and writer.

5 *The writer's energy and passion for the subject drive the writing, making the text lively, expressive, and engaging.*

- The tone and flavor of the piece fit the topic, purpose, and audience well.
- Clearly, the writing belongs to this writer and no other.
- The writer's sense of connection to the reader is evident.
- Narrative text is open, honest, and revealing.
- Expository or persuasive text is provocative, lively, and designed to prompt thinking and to hold a reader's attention.

3 *The writer seems sincere and willing to communicate with the reader on a functional, if somewhat distant level.*

- The writer has not quite found his or her voice but is experimenting—and the result is pleasant or intriguing, if not unique.
- Moments here and there amuse, surprise, or move the reader.
- The writer often seems reluctant to "let go" and thus holds individuality, passion, and spontaneity in check. The writer is "there"—then gone.
- Though clearly aware of an audience, the writer only occasionally speaks right to that audience or invites the audience "in."
- The writer often seems right on the verge of sharing something truly interesting—but then backs away as if thinking better of it.

1 *The writer seems somehow distanced from topic, audience, or both; as a result, the text may lack life, spirit, or energy. The writing reflects more than one of these problems—*

- The writer does not seem to reach out to the audience or to anticipate their interests and needs.
- Though it may communicate on a functional level, the writing takes no risks and does not involve or move the reader.
- The writer does not yet seem sufficiently at home with the topic to personalize it for the reader.

BEST COPY AVAILABLE

WORD CHOICE

Word choice is precision in the use of words—wordsmithery. It is the love of language, a passion for words, combined with a skill in choosing words that create just the mood, impression, or word picture the writer wants to instill in the heart and mind of the reader.

5 *Precise, vivid, natural language paints a strong, clear, and complete picture in the reader's mind.*

- The writer's message is remarkably clear and easy to interpret.
- Phrasing is original—even memorable—yet the language is never overdone.
- Lively verbs lend the writing power.
- Striking words or phrases linger in the writer's memory, often prompting connections, memories, reflective thoughts, or insights.

3 *The language communicates in a routine, workable manner; it gets the job done.*

- Most words are correct and adequate, even if not striking.
- Energetic verbs or memorable phrases occasionally strike a spark, leaving the reader hungry for more.
- Familiar words and phrases give the text an "old comfortable couch" kid of feel.
- Attempts at colorful language are full of promise, even when they lack restraint or control.

1 *The writer struggles with a limited vocabulary, searching for words or phrases to convey the intended meaning. The writing reflects more than one of these problems—*

- Vague words and phrases (She was *nice*. . . .*It* was *wonderful*. . . .The new budget *had impact*.) convey only the most general sorts of messages.
- Redundancy inhibits clarity and creativity.
- Clichés and tired phrases impair precision.
- Words are used incorrectly ("The bus *impelled* into the hotel.")
- The reader has trouble zeroing in on the writer's intended message.

BEST COPY AVAILABLE

SENTENCE FLUENCY

Sentence fluency is finely crafted construction combined with a sense of rhythm and grace. It is achieved through logic, creative phrasing, parallel construction, alliteration, absence of redundancy, variety in sentence length and structure, and a true effort to create language that literally cries out to be spoken aloud.

5 *An easy flow and rhythm combined with sentence sense and clarity make this text a delight to read aloud.*

- Sentences are well crafted, with a strong and varied structure that invites expressive oral reading.
- Purposeful sentence beginnings often show how a sentence relates to and builds on the one before it.
- The writing has cadence, as if the writer hears the beat in his or her head.
- Sentences vary in both structure and length, making the reading pleasant and natural, never monotonous.
- Fragments, if used, add to the style.

3 *The text hums along with a steady beat.*

- Sentences are grammatical and fairly easy to read aloud, given a little rehearsal.
- Some variation in length and structure enhances fluency.
- Some purposeful sentence beginnings aid the reader's interpretation of the text.
- Graceful, natural phrasing intermingles with more mechanical structure.

1 *A fair interpretive oral reading of this text takes practice. The writing reflects more than one of these problems—*

- Irregular or unusual word patterns make it hard to tell where one sentence ends and the next begins.
- Ideas are hooked together by numerous connectives (and . . . but . . . so then) to create one gangly, endless "sentence."
- Short, choppy sentences bump the reader through the text.
- Repetitive sentence patterns grow distracting or put the reader to sleep.
- Transitional phrases are either missing or so overdone they become distracting.
- The reader must often pause and reread to get the meaning.

BEST COPY AVAILABLE

CONVENTIONS

Almost anything a copy editor would attend to falls under the heading of conventions. This includes punctuation, spelling, grammar and usage, capitalization, and paragraphing—the spit-and-polish phase of preparing a document for publication. It does not (in this scoring guide) include layout, formatting, or handwriting.

5 *The writer shows excellent control over a wide range of standard writing conventions and uses them with accuracy and (when appropriate) creativity and style to enhance meaning.*

- Errors are so few and so minor that a reader can easily overlook them unless searching for them specifically.
- The text appears clean, edited, and polished.
- Older writers (grade 6 and up) create text of sufficient length and complexity to demonstrate control of conventions appropriate for their age and experience.
- The text is easy to mentally process; there is nothing to distract or confuse a reader.
- Only light touch-ups would be required to polish the text for publication.

3 *The writer shows reasonable control over the most widely used writing conventions and applies them with fair consistency to create text that is adequately readable.*

- There are enough errors to distract an attentive reader somewhat; however, errors do not seriously impair readability or obscure meaning.
- It is easy enough for an experienced reader to get through the text without stumbling, but the writing clearly needs polishing.
- Moderate editing would be required to get the text ready for publication.
- The paper reads much like a rough draft.

1 *The writer demonstrates limited control even over widely used writing conventions. The text reflects at least one of the following problems—*

- Errors are sufficiently frequent and/or serious as to be distracting; it is hard for the reader to focus on ideas, organization, or voice.
- The reader may need to read once to decode, then again to interpret and respond to the text.
- Extensive editing would be required to prepare the text for publication.

BEST COPY AVAILABLE

CRITERIA . . .

Language which helps us

1. Think about performance

2. Talk about performance

3. Distinguish among

STRONG,
DEVELOPING, &
JUST BEGINNING levels !

Rater Tips

Don't worry about . . .

- handwriting - assume this is as good as it gets
- titles - they are helpful, but not required
- length - score what is there (or not)
- topic - ideas come from lots of places!
- speed - some people just read faster/slower than others

Do worry about . . .

- the message - what do your scores tell the writer?
- the rubric - USE IT!
  (that's how we get consistency, reliability, accuracy)
- fatigue - take a stretch break
- skimming and scanning - writers can and will surprise us
- overreacting - is your pet peeve showing?

Common Pet Peeves

- ☹ **Beginning every sentence the same way**
- ☹ **Writing "ALOT" as one word**
- ☹ **Mixing "It's" and "ITS" and other common homonyms**
- ☹ **Messy handwriting**
- ☹ **Tiny handwriting**
- ☹ **Loopy handwriting**
- ☹ **Printing in full cap's**
- ☹ **Wordiness**
- ☹ **Phoniness**
- ☹ **Title doesn't match the text**

- ☹ Empty words used to "snow" the reader
- ☹ "THE END" at the end
- ☹ "Then I woke up and it was all a dream . . ." endings
- ☹ Blatant violence
- ☹ "CUZ"
- ☹ Paragraphs? What paragraphs?
- ☹ Safe, boring papers with no voice
- ☹ Broad generalizations that really say nothing
- ☹ Lack of originality
- ☹ Inconsistent use (or no use!) of capitals
- ☹ Beginning sentences with "BUT" or "AND"

Do these pet peeves sound familiar? Do you have one of your own? Write it here . . .

WATCH OUT. . . . . .

Following are COMMON sources of rater bias—any one of which can cause you to score unfairly:

- The positive-negative leniency error: A tendency to be too hard or too easy on everyone—just as a matter of principle.

- Appearance: Scoring up because the paper looks neat and presentable, or down because it looks messy. Judge content first. Good appearance IS important, but it is not part of the 6-trait scoring criteria.

- Length: Is longer better? No! In fact, length often works against a piece if there's too much "interruptive information". Papers that are too short, of course, cannot be scored fairly, but the real trick is to balance the need for good detail with an ability to be succinct.

- Fatigue: If you're never tired and bleary-eyed while you're doing this, you're either a machine, or you're sneaking your papers into someone else's stack. The point is, take an occasional break. You'll score faster (and MUCH more accurately) in the long run if you get up to stretch or have coffee/Coke/fresh air now and then.

- Personality clash: *I hate animal stories! I love sports papers! Oh, what a neat kid—he fishes with his dad! All this kid DOES is watch TV—he needs a couple 2's to wake him up to reality!* This is the don't-even-go-there approach to scoring. Try to be neutral. If you simply can't ("I *hate* iguanas, my father hated iguanas, and I'm *never* changing my mind"), give the paper to someone else. Think: What if it were your paper? Your child's?

- Skimming: You might think you know after the first 8 lines, but DO read the whole thing to be sure you're assessing the entire performance, not just the grand opening.

- Self-Scoring: Perceptive, intuitive readers fill in, anticipate, synthesize, and generally pull it all together for the writer. Be sure you're scoring the writer's work, not your skill in putting the puzzle together.

- Sympathy Score: *Her dog died. . . . She loves her grandpa so much. . . .*These situations tug at your heart, and rightly so. But the hard truth is, there are good and

not-so-good pet papers, grandparent papers, etc. Be sure you score the writing, not the circumstances.

- Pet Peeves, such as . . . .

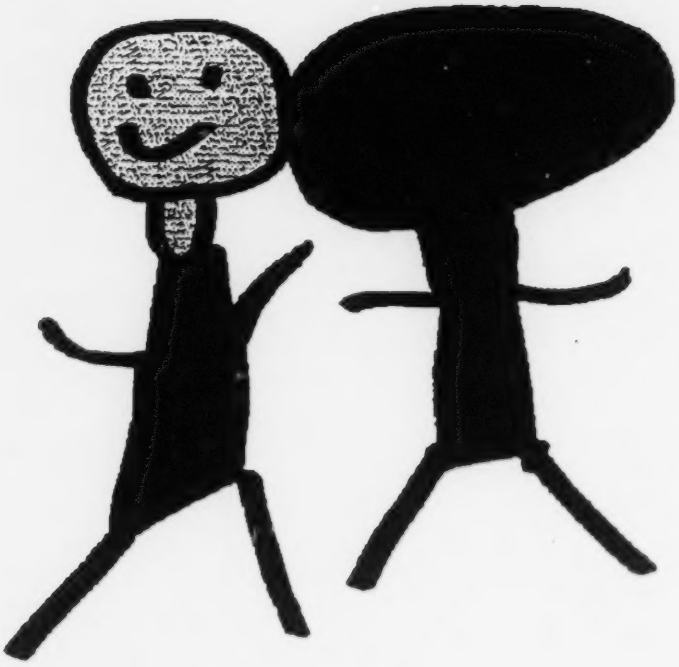
  - Big loopy writing
  - Teeny, tiny writing
  - Shifting tenses inappropriately
  - Writing in ALL CAPITALS
  - Tons! Of exclamation (!!) points!!! Yikes!!!! Wow!!!!!!
  - Messiness
  - Mixing *its* and *it's*
  - Run-on's
  - Mixing *are* and *our* or *their*, *there* and *they're*
  - The end (like I couldn't tell *that*)
  - The words and phrases *cool, awesome, rad, dude, neat, great, nice, far out, very, really*—or pick your own personal least favorites
  - The easy ending: "Then I woke up and it was all a dream"
  - Cliché adjectives: *fluffy* clouds, *crashing* waves
  - Total absence of paragraphs
  - Passionless writing
  - *Different than* instead of *Different from*
  - *Cuz*
  - *A lot*

And . . .watch out for pet peeves about which only you know. What are they??? (Yes, you do, even if you're a nice person.)

Good luck, best wishes, may you get many 5's and above all, on behalf of the student writers for whom your hard work makes such a difference, thank you.

SCORES FOR PRACTICE PAPERS

HOW DID YOU DO?



Northwest Regional Educational Laboratory
101 S.W. Main Street Portland, Or 97204

BEST COPY AVAILABLE

HOW DID I DO? 6-TRAIT PRACTICE PAPERS

Each of these 18 papers has been scored by several hundred (or more) classroom teachers who have been trained in the 6-trait model and who teach at or near the grade level of the writer. We have confidence in these scores, therefore, but we do NOT expect your scores to agree perfectly every time. Remember, the key word is <u>defensible</u>: can you point to language in the scoring guide that matches what you see in the student's performance. Then—allowing for some differences—here's a quick check on <u>your</u> performance as a rater:

1. CIRCLE any scores which match suggested scores exactly. Give yourself 2 POINTS for each circle.

2. Put a check ✓ beside each score that is within one point of the suggested score. If the suggested score is a 1, your score must be no higher than 2; if the suggested score is a 5, your score must be no lower than 4. Either a 2 or 4 will match a suggested score of 3, however! Give yourself 1 POINT for each adjacent score.

3. Put a big X through any score that does not match within one point: e.g., suggested score is 5, and you scored the paper a 3. Sorry—NO POINTS for scores two points or more apart.

4. Total your scores, and if your total is . . .

- <u>162 or more</u>—Wow! *Great* job. You know the traits well and are ready to score papers with confidence—and to help teach others.

- <u>132 or more</u>—Good job. Continue to read the scoring guide carefully and your scoring will grow even more accurate.

- <u>Under 132</u>—Please review the scoring guide one more time and rescore those papers that gave you trouble. Compare with a partner whose total for the 18 papers was higher than your own, and discuss any disagreements. You can do this—you just need to slow down or look a little more carefully!

BEST COPY AVAILABLE

Recommended Scores for Practice Papers

	Paper 1	Paper 2	Paper 3	Paper 4	Paper 5	Paper 6
Ideas	5	3	2	4	5	5
Organization	4	2	2	4	5	5
Voice	4	3	3	5	4	4
Word Choice	4	3	3	4	4	5
Fluency	4	2	3	4	5	5
Conventions	3	3	3	2	5	5